The Role of the Confluence Matrix in Exponential Tracking Convergence of Overparametrized Adaptive Feedforward Systems

D. S. Bayard[1]

Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive
Pasadena, CA 91109

Abstract

It is shown that persistent excitation (PE) conditions are overly stringent for ensuring exponential convergenceof the tracking error in overparametrized adaptive feedforward systems. Rather, only positive definiteness of a certain "confluence" matrix is required. The confluence matrix condition is easily satisfied and replaces the usual condition for positive definiteness of the "autocorrelation" matrix (i.e., the standard PE condition). This is important because the PE condition is unnecessarily stringent and essentially impossible to satisfy for overparametrized systems. It is concluded that the only penalties for overparametrization are that the optimal exponential rate of tracking error convergence is degraded by the condition number of the confluence matrix, and that the parameter errors converge exponentially on a reduced subspace rather than over the entire space. As a case study, the confluence matrix is examined in detail for the case of a Tap Delay-Line regressor with sinusoid excitation.

1 INTRODUCTION

In 1980, Bitmead and Anderson [8] proved that parameter convergence is *exponential* when persistent excitation (PE) conditions are satisfied in the adaptive gradient algorithm. One important consequence of exponential parameter convergence, is that the tracking error (which is typically linear in the parameter error) also converges exponentially. This relationship gives the (false) impression that exponential tracking error convergence requires the same stringent PE conditions as parameter convergence. Interestingly, there are several indications to the contrary. Using an approximate linear analysis, Glover [9] indicated as early as 1977 that exponential convergence of the tracking error is possible in the adaptive gradient algorithm with an overparametrized Tap Delay-Line regressor, and sinusoidal excitation, without any conditions on parameter convergence. More recently, Johansson [12] used a complete end-to-

[1] Mail Stop 198-326, bayard@bert2.jpl.nasa.gov

end Lyapunov analysis to demonstrate exponential tracking error convergence (to a bounded set) for a model reference adaptive control (MRAC) algorithm without persistent excitation or parameter convergence.

Motivated by Glover's results [9], this paper continues the investigation of exponential tracking error convergence in adaptive feedforward systems, with relaxed conditions on parameter convergence and persistent excitation. The paper extends Glover's results in the following directions,

1. The regressor basis can now be chosen arbitrary, compared to Glover's restriction to the Tap Delay-Line basis.
2. Glover's analysis rigorously requires that the number of taps goes to infinity (infinitely long regression vectors) to ensure exponential convergence. This was needed in his analysis to eliminate certain time-varying terms which prevented a purely linear time-invariant (LTI) stability argument based on pole locations of Laplace transforms. In the present analysis, the time-varying terms are part of the formulation, and the analysis is valid for the more realistic case of finite-length regressors.
3. Any excitation (with a certain PE reduction) can now be used compared to Glover's analysis which required sinusoidal excitation.
4. The present analysis is applicable to a broad class of adaptive feedforward algorithms while Glover's analysis was proved only for the LMS algorithm.

The first two extensions were also achieved in an earlier paper [2] at the cost of a somewhat restrictive condition on the regressor, i.e., by assuming that the regressor is a periodic function. This periodic assumption is not required in the present analysis.

A brief background is given in Section 2, where the confluence matrix is defined. The main results given in Section 3 show that if the confluence matrix is positive definite the adaptive feedforward operator $\mathcal{H}$ from error e to estimate $\hat{y}$ is input-output equivalent to an adaptive system with a PE regressor. This implies that tracking error convergence will be exponential for a large class of overparametrized adaptive feedforward systems without satisfying PE. In effect, *the condition for positive definiteness of the confluence matrix replaces the standard condition for positive definiteness of the "autocorrelation" matrix* (i.e., the well-known PE condition) found in textbooks [11][13][14]. This distinction is crucial for overparametrized adaptive systems which cannot satisfy PE, while they still may be designed to satisfy the weaker confluence matrix condition.

The main penalty incurred from overparametrizing comes from a reduction in the exponential rate of convergence compared to the full PE case. Specifically, it is shown in Section 4 for the adaptive gradient algorithm that the optimal exponential convergence rate degrades in proportion to the *condition number of the confluence matrix*. To give a sense for the amount of degradation, a case study is given in Section 5 analyzing adaptive systems with over-parametrized Tap Delay Line regressors and sinusoidal excitation. This special system was also studied by Glover, and allows his earlier results to be put into perspective. Conclusions are postponed until Section 6.

Good adaptive tracking performance is required in many applications where it is often not desirable or even possible to satisfy PE conditions [15]. The results in this paper indicate that the conditions for exponential tracking error convergence are considerably weaker than previously thought. The weakened condition implies that tracking error convergence is exponential for a large class of overparametrized adaptive feedforward systems with a wide range of possible regressor basis functions, excitations, and adaptation laws. This has many implications for the design of better adaptive feedforward systems in the future.

2 BACKGROUND

2.1 Adaptive Feedforward Systems

An estimate $\hat{y}$ of some signal y is to be constructed as a linear combination of the elements of a regressor vector $x(t) \in R^N$, i.e.,

Estimated Signal

$$\hat{y} = w(t)^T x(t) \tag{2.1}$$

where $w(t) \in RN$ is a parameter vector which is tuned in real-time using the adaptation algorithm,

Adaptation Algorithm

$$w = \mu \Gamma(p)[\tilde{x}(t) e(t)] \tag{2.2}$$

Here, $\Gamma(p)[\cdot]$ denotes the multivariable LTI transfer function I'(s). I where I'(s) is any S1S0 LTI transfer function in the Laplace s operator (the differentiation operator p will replace the Laplace operator s in all time-domain filtering expressions); the term $e(t) \in R^1$ is an error signal; $\mu > 0$ is an adaptation gain; and the signal $\tilde{x}$ is obtained by filtering the regressor x through any stable filter $F(p)$, *i.e.,*

Regressor Filtering

$$\tilde{x} = F(p)[x] \tag{2.3}$$

The notation $F(p)[\cdot]$ denotes the multivariable LTI transfer function F(s) $\cdot I$ with S1S0 filter $F(s)$, acting on the indicated vector time domain signal.

Equations (2. 1)-(2.3) taken together will be referred to as an *adaptive feedforward system*. Collectively, these equations define an important open-loop mapping from the error signal e to the estimated output $\hat{y}$. Because of its importance, the mapping from e to $\hat{y}$ will be denoted by the special character $\mathcal{H}$, i.e.,

$$\hat{y} = \mathcal{H}[e] \tag{2.4}$$

The special structure of $\mathcal{H}$ is depicted in Figure 2.1.

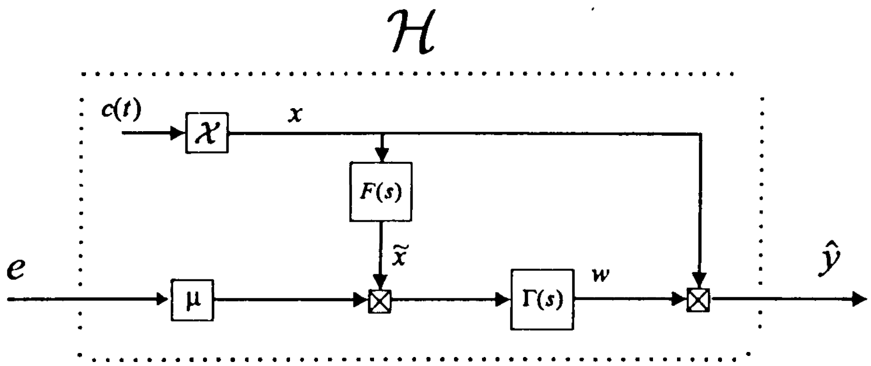

Figure 2.1: LTV operator $\hat{y} = \mathcal{H}[e]$ for adaptive system with regressor $\tilde{x}$, adaptation law $\Gamma(s)$, and regressor filter $F(s)$

REMARK 2.1 The definition of I’(s) is left intentionally general to include analysis of the gradient algorithm (i.e., with the choice I’(s) = 1/s), the gradient algorithm with leakage (i.e., $\Gamma(s)$ = 1/(s $+\sigma$); $\sigma \geq$ O), proportional-plus-integral adaptation (i.e., $\Gamma(s) = k_p + k_i/s$), or arbitrary linear adaptation algorithms of the designer’s choosing. Adaptation laws which are nonlinear or normalized (e.g., divided by the norm of the regressor), are not considered here since they do not have an equivalent LTI representation I’(s). ▪

REMARK 2.2 The use of the regressor filter $F(s)$ is (2.3) allows the unified treatment of many important adaptation algorithms including the well-known Filtered-X algorithm from the signal processing literature [17], and the Augmented Error algorithm of Monopoli (cf., [13]). ▪

2.2 Confluence Matrices and Overparametrization

Let $c(t) \in R^n$ be a bounded piecewise continuous signal vector, and let there exist positive constants β_1, β_2, T_o >0 such that,

$$\beta_1 \cdot I \leq \int_t^{t+T_o} c(\tau)c^T(\tau)d\tau \leq \beta_2 \cdot I \tag{2.5}$$

for all $t \geq$ O. Any signal $c(t)$ which satisfies these properties is said to be Persistently Exciting (PE) with bounds $\{\beta_1, \beta_2, T_o\}$ [11].

For the purpose of this paper, it will be assumed that the regressor $x(t) \in RN$ is linearly related to such a PE signal c(t) as follows,

$$x = \mathcal{X}c(t) \tag{2.6}$$

where $\mathcal{X} \in R^{N\times n}$. It is also assumed that $N \geq$ n in (2.6), so that $\mathcal{X}$ is a "tall" matrix and the adaptive feedforward system is *overparametrized.*

DEFINITION 2.1 *The matrix* $\mathcal{X}^T\mathcal{X}$ *is defined as the* confluence matrix *associated with a particular regressor x of the form (2.6).* ∎

The name "confluence matrix" has been chosen to reflect the fact that N signal channels seen at the output of the tall matrix $\mathcal{X}$ are effectively combined into a smaller number of n channels ($n \leq$ N) when forming $\mathcal{X}^T\mathcal{X}$. The confluence matrix will play an important role in characterizing the convergence properties of the overparametrized adaptive system.

Unfortunately, if the system is overparametrized the PE condition is impossible to satisfy. This will be shown by example.

EXAMPLE 2.1 Consider the case where,

$$c(t) = [\sin\omega_o t, \cos\omega_o t, \sin 2\omega_o t, \cos 2\omega_o t, \ldots \sin(m\omega_o t), \cos(m\omega_o t)]^T \tag{2.7}$$

Letting $T_O = 2\pi/\omega_o$ and using (2.7) one can calculate,

$$\int_t^{t+T_o} c(\tau)c^T(\tau)d\tau = \frac{T_o}{2}\cdot I \in R^{n\times n} \tag{2.8}$$

Hence, the PE condition (2.5) is satisfied with $\beta_1 = \beta_2 = T_o/2$.

Because of the form of $c(t)$ in (2.7), any regressor $x = \mathcal{X}c(t)$ will be periodic with period $T = 2\pi/\omega$. One can check the PE condition by computing the *autocorrelation matrix,*

$$\int_t^{t+T_o} x(\tau)x^T(\tau)d\tau = \mathcal{X}\int_t^{t+T_o} c(\tau)c^T(\tau)d\tau\mathcal{X}^T \tag{2.9}$$

$$= \frac{T_o}{2}\mathcal{X}\mathcal{X}^T \tag{2.10}$$

It is seen that the autocorrelation matrix is essentially the *outer-product* of the matrix X with itself. Consequently, if the problem is overparametrized (i.e., N > *n)* the matrix X is "tall", and it is impossible for $\mathcal{X}\mathcal{X}^T$ to be positive definite i.e., it is impossible to satisfy the PE conditions. ∎

It is a common belief that along with the loss of PE, comes the loss of exponential convergence. The main point of this paper is to show that this is not generally true, and in fact the exponential convergence properties of the tracking error e (and the parameters on a reduced subspace) do not depend on the *outer product condition* $XX^* > 0$, but rather depend on the *inner product condition* $X^TX > 0$. This result shifts the emphasis in the adaptive design from ensuring positive definiteness of the autocorrelation matrix, to ensuring positive definiteness of the confluence matrix. The main benefit is in the overparametrized case where the inner product condition is easy to satisfy, while the outer product condition is impossible to satisfy.

3 REGRESSOR REDUCTION TO PE

The next result shows that if the confluence matrix is positive definite, the adaptive feedforward operator $\mathcal{H}$ can always be reparametrized to have a PE regressor without changing its input-output properties.

THEOREM 3.1 (Regressor Reduction to PE) *Let the* **confluence** *matrix associated with the adaptive feedforward system (2.1)-(2.5') be positive definite,*

$$\mathcal{X}^T\mathcal{X} > 0 \tag{3.1}$$

Then,

(i) The input-output properties of the LTV operator $\mathcal{H}$ from e to $\hat{y}$ are invariant under the change of variables,

$$\eta(t) = \Lambda^{-\frac{1}{2}}P\mathcal{X}^T x(t) \tag{3.2}$$

$$p(t) = \Lambda^{-\frac{1}{2}}P\mathcal{X}^T w(t) \tag{3.3}$$

Here, $\eta \in R'$ and $p \in R^n$ are reduced-order regressor and parameter vectors, respectively, and matrices P, $\Lambda \in R^{n\times n}$ are defined from the eigenvalue decomposition of the confluence matrix,

$$\mathcal{X}^T\mathcal{X} = P^T\Lambda P \tag{3.4}$$

$$\Lambda = diag\{\lambda_1, ..., \lambda_n\} > 0 \tag{3.5}$$

where $P^T = P^{-1}$, and it is assumed that the eigenvalues are ordered as $\lambda_1 \geq \ldots \geq \lambda_n > 0$.

(ii) The reduced-order regressor $\eta \in R^n$ is PE with the bounds,

$$\beta_1\lambda_n \cdot I \leq \int_t^{t+T_o} \eta(\tau)\eta^T(\tau)d\tau \leq \beta_2\lambda_1 \cdot I \tag{3.6}$$

for all $t \geq 0$, where β_1, β_2, T_o are defined by the PE condition (2.5) for $c(t)$.

PROOF:

Proof of (i) The proof follows simply by the superposition and scaling properties of linear operators. As such, it can be proved graphically. Consider the sequence of block diagram rearrangements shown in Figure 3.1. Specifically, Figure 3.1 Part a. shows the initial adaptive system with overparametrized regressor x; Part b. shows the matrix X pushed through several scalar matrix blocks of the diagram; Part c. replaces the confluence matrix by its eigenvalue decomposition $\mathcal{X}^T\mathcal{X} = P^T\Lambda P$; Part d. pushes the matrix factor $\Lambda^{\frac{1}{2}}P$ back through several scalar matrix blocks. The resulting block diagram is driven by the regressor η which related to c by the nonsingular transformation $\eta = \Lambda^{\frac{1}{2}}Pc$, and hence is PE.

Proof of (ii) Define,

$$M = \int_t^{t+T_o} \eta(\tau)\eta^T(\tau)d\tau \tag{3.7}$$

Substituting η = **A*Pc** into (3,7), gives

$$M = \Lambda^{\frac{1}{2}} P \int_{t}^{t+T_o} c(\tau)c^T(\tau)d\tau P^T \Lambda^{\frac{1}{2}} \tag{3.8}$$

Using the PE property of $c(t)$ in (2.5) and (3.5) gives,

$$\overline{\sigma}(M) \leq \beta_2 \lambda_1 \tag{3.9}$$

$$\underline{\sigma}(M) \geq \beta_1 \lambda_n \tag{3.10}$$

which is equivalent to (3.6) as desired. ∎

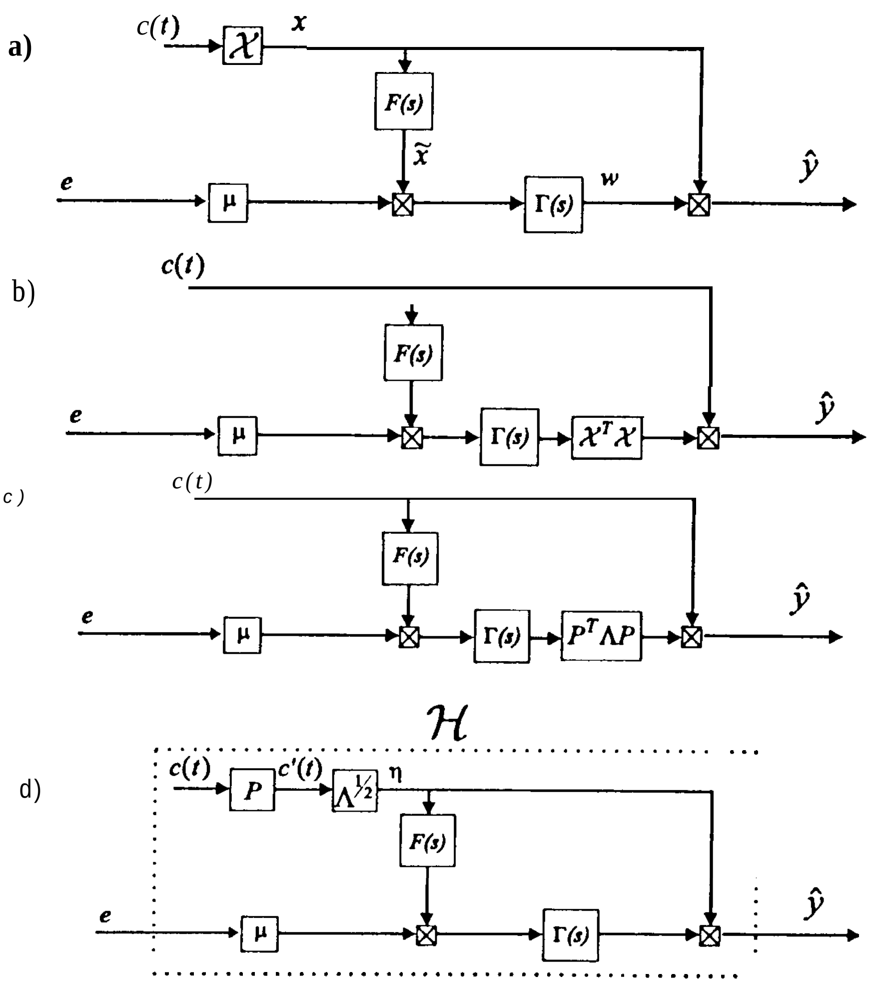

Figure 3.1: Proof of Theorem 3.1 by Block diagram rearrangements

In order to use the result of Theorem 3.1, the overparametrized regressor $x \in R^N$ must satisfy the decomposition $x(t) = X_1 c_1(t)$ for some matrix $\mathcal{X}_1 \in R^{N \times n}$, and some PE signal vector $c_1 \in R^n$. Fortunately, in many applications such a decomposition is straightforward to find. Consider the next example.

EXAMPLE 3.1 (Harmonic Regressor)

It is common for the regressor x to have a harmonic form, i.e., where its elements are composed of linear combinations of m sinusoidal frequencies $\{\omega_i\}_{i=1}^m$. For this regressor, the signal c_1 can be chosen of length $2m$ having elements,

$$c_1(t) = [\sin\omega_1 t, \cos\omega_1 t, \ldots, \sin\omega_m t, \cos\omega_m t]^T \in R^{2m} \tag{3.11}$$

This construction guarantees the existence of a matrix $\mathcal{X}_1 \in R^{2m}$ in the desired decomposition $x = \mathcal{X}_1 c_1(t)$. ∎

Given matrix $\mathcal{X}_1$ and PE signal c_1 such that z = $\mathcal{X}_1 c_1(t)$, the confluence matrix $\mathcal{X}_1^T \mathcal{X}_1$ may or may not be positive definite. Remarkably, it turns out (neglecting the trivial case where $\mathcal{X}$ = O), that the confluence matrix *can always be made positive definite by transforming c_1 to a reduced-order C_2*. To see this, consider the singular-value decomposition of Xl,

$$\mathcal{X}_1 = U\Sigma V^T \tag{3.12}$$

where the matrices $U \in R^{N \times 2m}$, $\Sigma \in R^{2m\times 2m}$ and $V \in R^{2m\times 2m}$ are partitioned compatibly as, $U = [U_a | U_b], \Sigma = diag\{\Sigma_a, \Sigma_b\}$ and such that if the confluence matrix is singular then $\Sigma_a > 0$ and $\Sigma_b = 0$. Define $\mathcal{X}_2 = U_a$ and extract C_2 from the equality $[c_2(t)|0]^T = \Sigma_a V^T c_1(t)$. Then x = $X_2 c_2(t)$ and $\mathcal{X}_2^T \mathcal{X}_2$ >0 as desired.

In summary, if a regressor is representable in the form $x = \mathcal{X}_1 c_1(t)$ for some PE signal $c(t)$ then it either it has the positive definite confluence matrix $\mathcal{X}_1^T \mathcal{X}_1 > 0$, or by reduction of c_1 to C_2 is transformable to the form x = $\mathcal{X}_2 c_2(t)$ with PE signal $c_2(t)$, which has the positive definite confluence matrix $\mathcal{X}_2^T \mathcal{X}_2$ >0. In either case, the confluence matrix is positive definite and the results of Theorem 3.1 indicate that the operator $\mathcal{H}$ can be reparametrized to have a PE regressor.

REMARK 3.1 When the regressor has the harmonic form x = $\mathcal{X}c$ for c of the form (3.1 1), it sometimes turns out that the mapping $\mathcal{H}$ is *purely LTI*. This remarkable property considerably simplifies the analysis, and lies at the heart of Glover's approach in [9]. It has been found recently [3] [4] that this LTI property occurs *if and only if the confluence matrix has a pairwise diagonal structure,* i.e.,

$$\mathcal{X}^T \mathcal{X} = D^2 \tag{3.13}$$

where,

$$D^2 = diag[d_1^2, d_1^2, d_2^2, d_2^2, \ldots, d_m^2, d_m^2] \tag{3.14}$$

This property also underscores the importance of the confluence matrix for characterizing overparametrized adaptive systems. ∎

The input-output properties of $\mathcal{H}$ play a critical role in determining the convergence properties of the tracking error e in closed-loop. The result of Theorem 3.1 is important becuase it shows that the input-output properties of the operator $\mathcal{H}$ (even overparametrized) are identical to one which is reparametrized to have a (reduced-order) PE regressor. Given such a PE regressor, many proofs of exponential stability and robustness exist in the literature (cf., [11][13][14]). As applied to the present case, these proofs ensure exponential convergence of the tracking error (and the parameters on a reduced subspace) in closed-loop, and retention of certain robustness properties which are invariant under the internal reparametrization of $\mathcal{H}$ (e.g., such as boundedness with respect to additive bounded noise disturbances in e).

It will not be possible to delineate all possible ways to prove exponential stability given a PE regressor, since the proofs are algorithm specific. Instead, the adaptive gradient algorithm is chosen in the next section as a simple and representative algorithm to examine these exponential convergence properties in more det ai 1.

4 EXPONENTIAL CONVERGENCE

In this section, exponential convergence with overparametrization is examined for the adaptive gradient algorithm.

4.1 Adaptive Gradient Algorithm

Let the $y(t) \in R^1$ and $x(t) \in RN$, be known signals and assume there exists a constant parameter vector $w^o \in RN$ such that,

$$y(t) = w^{oT} x(t) \tag{4.1}$$

for all t >0. Uniqueness of w^o is not required (i.e., the system can be overparametrized). An estimate $\hat{y}$ of y is constructed as,

$$\hat{y} = w(t)^T x(t) \tag{4.2}$$

where $w(t)$ is tuned in real-time using the adaptive gradient algorithm [13] (i.e., set $\Gamma(s)$ = 1/s in (2.2) and $F(s)$ = 1 in (2.3)).,

$$\dot{w} = \mu x(t) e(t) \tag{4.3}$$

with adaptation gain μ >0. The tracking error is defined as,

$$e(t) = y(t) - \hat{y}(t) \tag{4.4}$$

and the parameter error is defined as,

$$\phi(t) = w^o - w(t) \tag{4.5}$$

Using (4.1)(4.2)(4.4)(4.5), the tracking and parameter errors can be related as follows,

$$e = \phi^T x(t) \tag{4.6}$$

Assuming that the true parameter w^o does not vary with 'time, (i.e., $\dot{w}^o = 0$), it follows from (4.3)(4.5) that,

$$\dot{\phi} = \dot{w}^o - \dot{w} = -\mu x e = -\mu x x^T \phi \tag{4.7}$$

'I'his equation characterizes the propagation of the parameter error.

4.2 Exponential Convergence Properties

It is convenient at this point to review a well-known stability argument. Define the Lyapunov function candidate,

$$V = \frac{1}{2}\phi^T \phi \tag{4.8}$$

Taking the derivative of (4.8) and using (4.1)-(4,7) yields,

$$\dot{V} = -\mu e \phi^T x = -\mu e^2 \leq 0 \tag{4.9}$$

This proves that ϕ remains bounded. If x is bounded, then from (4.6) the error e remains bounded. Furthermore, if x is bounded, then $\ddot{V}$ is bounded, $\dot{V}$ is uniformly continuous, and Barbalat's lemma ([13], pg. 85, and 276), can be applied to ensure that $\lim_{t\to\infty} e = 0$. This well known argument ensures that the error converges to zero as desired.

While the above argument ensures that e converges to zero, it does not indicate *how fast* it converges. Additional conditions such as persistent excitation are typically imposed which ensure exponential convergence of e to zero.

Persistent excitation conditions in adaptive algorithms have been studied by many researchers. Early results can be found in Astrom and Bohlin [1] where the PE condition is expressed in terms of positive definiteness of the autocorrelation function formed from the regressor. Subsequently, Bitmead and Anderson [8] proved that parameter convergence is *exponential* when PE conditions are satisfied in the adaptive gradient algorithm and the normalized adaptive gradient algorithms. Explicit upper and lower bounds on the exponential response can be found in [16]. A general discussion of the PE condition is given in [7] and an effort to unify many definitions can be found in [18].

As an example, consider the case without overparametrization (i.e., N = n), so that $x(t) \in R^n$ is bounded and PE satisfying, say,

$$||x(t)|| \leq \bar{x} < \infty; \text{ for all } t \geq 0 \tag{4.10}$$

$$\alpha_1 \cdot I \leq \int_t^{t+\delta} x(\tau)x^T(\tau)d\tau \leq \alpha_2 \cdot I \tag{4.11}$$

for some $\alpha_1, \alpha_2, \delta > 0$. Then it is well known (cf., [13][14][11]), that the error e converges exponentially. Specifically, there exist constants $p. \geq O$, $\alpha > 0$ such that,

$$|e| \leq p_0 e^{-\alpha t} \tag{4.12}$$

The precise expression for α is given in Lemma A. 1 of Appendix A as,

$$\alpha = \frac{1}{2\delta} \ln \left(\frac{1}{1 - \alpha_3} \right) \tag{4.13}$$

$$p_0 = \left(\frac{1}{1 - \alpha_3} \right)^{\frac{1}{2}} \cdot ||\phi(0)|| \tag{4.14}$$

$$\alpha_3 = \frac{2\mu\alpha_1}{(1 + \mu\alpha_2\sqrt{n})^2} \tag{4.15}$$

The convergence rate α in (4.13) is a function of μ through the expression (4.15). For small μ the rate can be approximated by,

$$\alpha \simeq \frac{\mu\alpha}{\delta} \tag{4.16}$$

The fastest convergence rate is found by optimizing α in (4.13) with respect top. Specifically, the condition $d\alpha/d\mu = O$ can be solved to give the optimal gain as,

$$\mu^* = \frac{1}{\alpha_2\sqrt{n}} \tag{4.17}$$

Substituting (4.17) into (4.13) gives,

$$\alpha^* = \frac{1}{2\delta} \ln \left(\frac{1}{1 - \frac{1}{2\sqrt{n}} \cdot \frac{\alpha_1}{\alpha_2}} \right) \tag{4.18}$$

It is seen that *the optimized rate α^* improves monotonically with the ratio α_1/α_2.* This ratio is precisely the reciprocal condition number of the autocorrelation matrix (4.11), and motivates keeping this condition number as close to unity as possible for fast convergence (assuming it is optimally tuned with p“).

Exponential convergence for the overparametrized adaptive gradient algorithm is examined next.

THEOREM 4.1 (Overparametrized Adaptive Gradient) *Assume there exists a $w^o \in RN$ such that (4.1) holds for all $t \geq O$, and that the adaptive gradient algorithm (4.2)-(4.7) is used to tune w, giving the following error system,*

$$e = \phi^T x \tag{4.19}$$

$$\dot{\phi} = -\mu x x^T \phi \tag{4.20}$$

Let $c(t) \in R^n$ *be a bounded piecewise continuous signal vector which is PE, i.e., let there exist positive constants* $\beta_1, \beta_2, T_O > 0$ *such that,*

$$\beta_1 \cdot I \leq \int_t^{t+T_o} c(\tau)c(\tau)^T d\tau \leq \beta_2 \cdot I; \; for\ all\ t \geq 0 \tag{4.21}$$

$$||c(t)|| \leq \bar{c} < \infty; \; for\ all\ t \geq 0 \tag{4.22}$$

Let the regressor $x(t) \in R^N$ *be linearly related to the PE signal* $c(t)$ *as follows,*

$$x = \mathcal{X} c(t) \tag{4.23}$$

where $\mathcal{X} \in R^{N \times n}$ *and* $N \geq n$ *(i. e., the system can be overparametrized). Let the confluence matrix be positive definite,*

$$\mathcal{X}^T \mathcal{X} > 0 \tag{4.24}$$

and let the eigenvalue decomposition of the confluence matrix be given as,

$$\mathcal{X}^T \mathcal{X} = P^T \Lambda P \tag{4.25}$$

$$\Lambda = diag\{\lambda_1, ..., \lambda_n\} > 0 \tag{4.26}$$

where $P^T = P^{-1}$, *and* $\lambda_1 \geq \ldots \geq \lambda_n > 0$.

Then,

(i) *The error system (4.19)(4.20) can be written equivalently as the reduced system,*

$$e = r^T \eta \tag{4.27}$$

$$\dot{r} = -\mu \eta \eta^T r \tag{4.28}$$

where the reduced regressor $\eta \in R^n$ *and parameter error* $r \in R^n$ *are given, respectively, by,*

$$\eta = \Lambda^{\frac{1}{2}} P c(t) \tag{4.29}$$

$$r = \Lambda^{-\frac{1}{2}} P \mathcal{X}^T \phi \tag{4.30}$$

(ii) *The reduced regressor* $\eta \in R^n$ *is PE with the bounds,*

$$\beta_1 \lambda_n \cdot I \leq \int_t^{t+T_o} \eta(\tau)\eta^T(\tau) d\tau \leq \beta_2 \lambda_1 \cdot I \tag{4.31}$$

for all $t \geq 0$.

(iii) *The tracking error* e *and reduced parameter error* r *converge to zero exponentially as,*

$$||r|| \leq p_0 e^{-\alpha t} \tag{4.32}$$

$$|e| \leq p_0 \lambda_1^{\frac{1}{2}} \bar{c} e^{-\alpha t} \tag{4.33}$$

where,

$$\alpha = \frac{1}{2T_o} \ln\left(\frac{1}{1-\alpha_3}\right) \tag{4.34}$$

$$p_0 = \left(\frac{1}{1-\alpha_3}\right)^{\frac{1}{2}} \cdot ||r(0)|| \tag{4.35}$$

$$\alpha_3 = \frac{2\mu\beta_1\lambda_n}{(1+\mu\beta_2\lambda_1\sqrt{n})^2} \tag{4.36}$$

Letting μ *be sufficiently small (i. e., such that* $\mu << 1/(\beta_2\lambda_1 T_o\sqrt{n})$*),* gives,

$$\alpha \simeq \mu\beta_1\lambda_n/T_o \tag{4.37}$$

$$p_o \simeq (1+\mu\beta_1\lambda_n)||r(0)|| \tag{4.38}$$

PROOF:

Proof of (i): Using the transformed vectors η and r, the error equation (4.19) can be written as follows,

$$\begin{aligned} e &= \phi^T x = \phi^T \mathcal{X} c(t) && (4.39)\\ &= \phi^T \mathcal{X} P^T \Lambda^{-\frac{1}{2}} \Lambda^{\frac{1}{2}} P c(t) && (4.40)\\ &= r^T \eta && (4.41) \end{aligned}$$

which is (4.27) as desired. Likewise, for the adaptation law (4.20) one has,

$$\dot{\phi} = -\mu x x^T \phi = -\mu \mathcal{X} c c^T \mathcal{X}^T \phi \tag{4.42}$$

Multiplying both sides of (4.42) on the left by $\Lambda^{-\frac{1}{2}} P \mathcal{X}^T$ gives,

$$\begin{aligned} \Lambda^{-\frac{1}{2}} P \mathcal{X}^T \dot{\phi} &= -\mu \Lambda^{-\frac{1}{2}} P \mathcal{X}^T \mathcal{X} c c^T \mathcal{X}^T \phi && (4.43)\\ &= -\mu \Lambda^{-\frac{1}{2}} P P^T \Lambda P c c^T \mathcal{X}^T \phi && (4.44)\\ &= -\mu \Lambda^{\frac{1}{2}} P c c^T \mathcal{X}^T \phi && (4.45)\\ &= -\mu \Lambda^{\frac{1}{2}} P c c^T P^T \Lambda^{\frac{1}{2}} \Lambda^{-\frac{1}{2}} P \mathcal{X}^T \phi && (4.46) \end{aligned}$$

Substituting (4.29) and (4.30) into both sides of (4.46) gives,

$$\dot{r} = -\mu \eta \eta^T r \tag{4.47}$$

which is (4.28) as desired.

Proof of (ii): Identical to Proof of (ii) of Theorem 3.1.

Proof of (iii) By Part (ii) the reduced regressor η in *(4.29) is* persistently exciting. It follows from Lemma A. 1 of Appendix A that the reduced error system (4.27)(4.28) converges exponentially. In light of the PE bounds in (4.31), Lemma A.1 can be applied with $\alpha_1 = \beta_1 \lambda_n$, $\alpha_2 = \beta_2 \lambda_1, \delta = T_o$ and $\bar{x} = \lambda_1^{\frac{1}{2}} \bar{c}$ to give results (4.32)-(4.38) as desired. ■

4.3 Discussion

Intuitively, the persistent excitation conditions are eliminated in Theorem 3.1 by avoiding the need for convergence of the full parameter vector w in the proof. Rather, the '(degree" to which the given regressor x is persistently exciting is indicated by the size *n* of the vector $c(t)$. The parameter error vector ϕ is transformed to become the smaller vector $r = \Lambda^{-\frac{1}{2}} P \mathcal{X}^T \phi$ where $r \in R^n$ is defined on a subspace which is excited persistently. Since the regressor η associated with r is persistently exciting, the reduced error vector r converges exponentially, which from (4.27) ensures exponential convergence of e.

In light of 4.1 it reasonable to examine the penalties incurred from overparametrization. As mentioned above, the first issue is that the parameters converge on a subspace rather than over the entire space. However, this is only a penalty if the adaptive system is being used for parameter estimation properties. In fact, for parameter estimation there is often no other choice but to use a minimal parametrization or to increase the richness of the excitation. In contrast, full parameter convergence is not an issue in applications which use the adaptive feedforward system for its tracking properties ,i. e, where convergence of e to zero is the main consideration.

The second penalty, is that the exponential rate of convergence is degraded. This effect will be examined more closely. If one constructs the ideal regressor as $x(t) = c(t)$ then the regressor is PE with parameters $\{\beta_1, /3_2, T_o\}$ and one would achieve an optimal exponential convergence of (set $(\alpha_1 = \beta_1, \alpha_2 = /?_2, \delta = T_o$ in (4.18)),

$$\alpha^* = \frac{1}{2\delta} \ln \left(\frac{1}{1 - \frac{1}{2\sqrt{n}} \cdot \frac{\beta_1}{\beta_2}} \right) \tag{4.48}$$

However, if the regressor is overparametrized as $x = \mathcal{X} c(t)$, then Theorem 4.1 part (ii) indicates that the PE parameters of the equivalent reduced-order regressor η are degraded to $\beta_1 \lambda_n, \beta_2 \lambda_1, T_o$. This modifies the optimal convergence rate to,

$$a" = \frac{1}{2\delta} \text{i n} \left(\frac{1}{1 - \frac{1}{2\sqrt{n}} \cdot \frac{\beta_1}{\beta_2} \cdot \frac{\lambda_n}{\lambda_1}} \right) \tag{4.49}$$

i.e., the rate is degraded by (a monotonic function of) the ratio λ_n / λ_1. This ratio is precisely the reciprocal of the condition number of the confluence matrix $\mathcal{X}^T\mathcal{X}$, denoted as,

$$\kappa(\mathcal{X}^T\mathcal{X}) \triangleq \frac{\lambda_1}{\lambda_n} \tag{4.50}$$

In words then, *the optimal exponential convergence rate for an overparametrized adaptive system degrades (compared to the full PE case) in monotonically with the condition number of the confluence matrix.*

The reader is warned that the condition number (4.50) can be quite large. *In fact, most of the bad experiences that researchers have with sluggish convergence of overparametrized adaptive systems can be traced to this quantity.* The usual tendency is to blame overparametrization and lack of PE. However, the main point to be made here is that, on the contrary, overparametrized systems have exponential convergence just like full PE systems, and can provide good performance if care is taken to ensure that the condition number (4.50) is well behaved.

To better assess the important issue of convergence rate degradation due to overparametrization, the condition number of the confluence matrix will be examined in more detail in the next section on a specific problem of practical interest.

5 SINUSOIDAL TDL REGRESSORS

In this section, the confluence matrix condition number is explicitly bounded in the case of a Tap Delay-Line (TDL) regressor with sinusoidal excitation. The main reason is to evaluate the amount of rate degradation expected from overparametrization in a system which is used quite often in practice. This is precisely the adaptive system studied by Glover in [9], and the results can be directly compared.

5.1 Bounded Tone Sets

DEFINITION 5.1 *Given time delay T and spacing parameter $0 < \underline{\nu} < \pi/2$, a Bounded Tone Set $\Omega(m, T, \underline{\nu})$ is defined as any set of m frequencies $\{\omega_i\}_{i=1}^m$, such that,*

$$\Omega(m,T,\underline{\nu}) \triangleq \left\{ \begin{array}{ll} \{\omega_i\}_{i=1}^m : & 0 < \underline{\nu} < \pi/2; \\ 0 < \underline{\nu} < \omega_i T \leq \pi - \underline{\nu} & \text{for all } i = 1, \ldots, m; \\ |\omega_i - \omega_j| T \geq 2\underline{\nu} & \text{for all } i \neq j \end{array} \right\} \tag{5.1}$$

■

Simply stated, a Bounded Tone Set is a set of frequencies $\{\omega_i\}_{i=1}^{m}$ which are bounded away from O, π/T and each other. The definition is not very restrictive since any signal comprised of a finite number of distinct sinusoids lies in a Bounded Tone Set when T is chosen sufficiently small (i.e., to ensure Nyquist sampling of its highest component). The essential need for Definition 5.1 is to define the *minimum tone spacing* parameter $\underline{\nu}$ which will play a critical role in subsequent results.

5.2 TDL Background

Let the components of the regressor $x = [x_1, ..., x_N]^T \in RN$ be defined by filtering a signal $\xi(t) \in R^1$ through a Tap Delay-Line with N taps and tap delay T, i.e.,

$$x_\ell = e^{-(\ell-1)pT}\xi, \quad \ell = 1, ..., N \tag{5.2}$$

where the measured signal ξ is given by the following sum of m sinusoids,

$$\xi(t) = \sum_{i=1}^{m} C_i \sin(\omega_i t + \theta_i) \tag{5.3}$$

$$= \sum_{i=1}^{m} a_{i1}\sin(\omega_i t) + a_{i2}\cos(\omega_i t) \tag{5.4}$$

and where ail = $C_i \cos \theta_i$ and $a_{i2} = C_i \sin \theta_i$. Assume that the signal ξ in (5.4) is comprised of frequencies ω_i lying in the bounded tone set,

$$\{\omega_i\} \in \Omega(m, T, \underline{\nu}) \tag{5.5}$$

for some O $< \underline{\nu} < \pi/2$.

It is shown in [3] that the TDL regressor $x(t)$ in (5.2) can be written in form,

$$x = \mathcal{X}c(t) \tag{5.6}$$

$$c(t) = [\sin(\text{oIi}), \cos(\omega_1 t), ... \sin(\omega_m t), \cos(\omega_m t)] \in R^{2m} \tag{5.7}$$

where the matrix X' $\in R^{Nx2m}$ satisfies,

$$X = Q A \tag{5.8}$$

$$Q = [S_1, C_1, ..., S_m, \boldsymbol{Cm}] \in R^{Nx2m} \tag{5.9}$$

$$S_i = \left| \begin{matrix} 1 \\ \sin(\omega_i T) \\ \vdots \\ \sin((N-1)\omega_i T) \end{matrix} \right| \text{, RN;} \quad C_i = \begin{bmatrix} 1 \\ \cos(\omega_i T) \\ \vdots \\ \cos((N-1)\omega_i T) \end{bmatrix}^T \in R^N \tag{5.10}$$

$$\mathcal{A} = \begin{bmatrix} A_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & A_m \end{bmatrix} \in R^{2m\times 2m} \tag{5.11}$$

$$A_i = \begin{bmatrix} a_{i2} & -a_{i1} \\ a_{i1} & a_{i2} \end{bmatrix} \in R^{2\times 2}; \quad i = 1, ..., m \tag{5.12}$$

The following result istakenfrom [3] without proof.

LEMMA 5.1 *Let $Q, \underline{\nu}$ and m be as defined in (5.2)-(5.11). Then,*

$$\overline{\sigma}(Q^T Q - \frac{N}{2} \cdot I) \leq \frac{m\pi}{2\underline{\nu}} \tag{5.13}$$

PROOF: see Theorem 7.1 and Appendix B of [3]. ∎

5.3 Properties of Confluence Matrix for TDL Regressors

The desired properties of the confluence matrix are given in the next result.

THEOREM 5.1 *Let the components of the regressor $x = [x_1, \ldots, x_N]^T \in R^N$ be defined by filtering a signal $\xi(t) \in R^1$ through a Tap Delay-Line with N taps and tap delay T, i.e.,*

$$x_\ell = e^{-(\ell-1)pT}\xi, \; \ell = 1, \ldots, N \tag{5.14}$$

where the measured signal ξ is given by the following sum of m sinusoids,

$$\xi(t) = \sum_{i=1}^{m} C_i \sin(\omega_i t + \theta_i) \tag{5.15}$$

and frequencies $\{\omega_i\}_{i=1}^m$ lie in a bounded tone set $\Omega(m, T, \underline{\nu})$. Let the number of taps N be sufficiently large to satisfy,

$$N > \frac{m\pi}{\underline{\nu}} \tag{5.16}$$

Then, the maximum eigenvalue λ_1, minimum eigenvalue λ_{2m}, and condition number κ of the confluence matrix $X^T X$ for the TDL regressor (5.6)-(5.12) are bounded as follows,

$$\lambda_1(\mathcal{X}^T\mathcal{X}) = \overline{\sigma}(\mathcal{X}^T\mathcal{X}) \leq C_{max}^2 \cdot (\frac{N}{2} + \frac{m\pi}{2\underline{\nu}}) \tag{5.17}$$

$$\lambda_{2m}(\mathcal{X}^T\mathcal{X}) = \underline{\sigma}(\mathcal{X}^T\mathcal{X}) \geq C_{min}^2 \cdot (\frac{N}{2} - \frac{m\pi}{2\underline{\nu}}) \tag{5.18}$$

$$\kappa(\mathcal{X}^T\mathcal{X}) \leq \frac{C_{max}^2}{C_{min}^2} \cdot \frac{(N + \frac{m\pi}{\underline{\nu}})}{(N - \frac{m\pi}{\underline{\nu}})} \tag{5.19}$$

where $C_{max} \triangleq \max_i\{C_i\}$ *and* $C_{min} \triangleq \min_i\{C_i\}$

PROOF:

Equation (5.17) is proved as follows,

$$\overline{\sigma}(\mathcal{X}^T\mathcal{X}) = \overline{\sigma}(\mathcal{A}^TQ^TQ\mathcal{A}) \tag{5.20}$$

$$\leq \overline{\sigma}^2(\mathcal{A})\overline{\sigma}(Q^TQ) \tag{5.21}$$

$$= C_{max}^2 \cdot \overline{\sigma}(Q^TQ) \tag{5.22}$$

$$= C_{max}^2 \cdot \overline{\sigma}(\frac{N}{2} \cdot I + Q^TQ - \frac{N}{2} \cdot I) \tag{5.23}$$

$$\leq C_{max}^2 \cdot \left(\overline{\sigma}(\frac{N}{2} \cdot I) + \overline{\sigma}(Q^TQ - \frac{N}{2} \cdot I)\right) \tag{5.24}$$

$$\leq C_{max}^2 \cdot (\frac{N}{2} + \frac{m\pi}{2\underline{\nu}}) \tag{5.25}$$

Here, equation (5.20) follows by substituting (5.8); equation (5.21) follows by properties of singular values; equation (5.22) follows from the structure of $\mathcal{A}$ in (5.11) and (5.12) combined with the fact that $A_i^TA_i = C_i^2 \cdot 1_{22}$ (and hence $\overline{\sigma}(A_i) = \underline{\sigma}(A_i) = C_i$), and the definition of C_{max}, C_{min}; equation (5.23) follows by adding and subtracting the term $\frac{N}{2} \cdot I$; equation (5.24) follows by a property of singular values; and equation (5.26) follows by Lemma 5.1, which is the desired result (5.25).

Result (5. 17) is proved by the following similar sequence of steps,

$$\underline{\sigma}(\mathcal{X}^T\mathcal{X}) = \underline{\sigma}(\mathcal{A}^TQ^TQ\mathcal{A}) \tag{5.26}$$

$$\geq \underline{\sigma}^2(\mathcal{A})\underline{\sigma}(Q^TQ) \tag{5.27}$$

$$= C_{min}^2 \cdot \underline{\sigma}(Q^TQ) \tag{5.28}$$

$$= C_{min}^2 \cdot \underline{\sigma}(\frac{N}{2} \cdot I + Q^TQ - \frac{N}{2} \cdot I) \tag{5.29}$$

$$\geq C_{min}^2 \cdot \underline{\sigma}(\frac{N}{2} \cdot I) - \overline{\sigma}(Q^TQ - \frac{N}{2} \cdot I) \tag{5.30}$$

$$\geq C_{min}^2 \cdot (\frac{N}{2} - \quad) \tag{5.31}$$

Result (5.19) is proved by using results (5.17) and (5.18) as follows,

$$\kappa(\mathcal{X}^T\mathcal{X}) \triangleq \frac{\overline{\sigma}(\mathcal{X}^T\mathcal{X})}{\underline{\sigma}(\mathcal{X}^T\mathcal{X})} \tag{5.32}$$

$$\leq \frac{C_{max}^2}{C_{min}^2} \cdot \frac{(\frac{N}{2} + \frac{m\pi}{2\nu})}{(\frac{N}{2} - \frac{m\pi}{2\nu})} \tag{5.33}$$

$$= \frac{C_{max}^2}{C_{min}^2} \cdot \frac{(N + \frac{m\pi}{\nu})}{(N - \frac{m\pi}{\nu})} \tag{5.34}$$

It has been tacetly assumed in (5.33)-(5.34) that the denominators will not be negative. This can be enforced rigorously by choosing the number of taps sufficiently large to satisfy $N > \frac{m\pi}{\nu}$, which is the stated condition (5.16) on N. ■

The results of Theorem 5.1 indicate that,

1. As the number of taps N becomes large, (5.19) indicates that the condition number of the confluence matrix is bounded by the worst-case tone amplitude ratio $\frac{C_{max}^2}{C_{min}^2}$. This is essentially the limiting penalty for TDL overparametrization with a large number of taps, with regard to degrading the rate of exponential convergence of the tracking error.

2. Since the limit $\frac{C_{max}^2}{C_{min}^2}$ is achieved quickly as N begins to dominate $\frac{m\pi}{\nu}$ in (5.19), one approaches "diminishing returns" and there is no practical reason to increase the number of taps in the TDL to, say, larger than $N = 10\frac{m\pi}{\nu}$. This gives a useful guideline for choosing the number of taps in the overparametrized TDL (this can otherwise be a very confusing choice in practice).

3. For the TDL regressor with sinusoidal excitation, it has been shown in [6][3] that,

$$\frac{1}{N}\mathcal{X}^T\mathcal{X} = D^2 + \Delta \tag{5.35}$$

where D^2 is pairwise diagonal of the form,

$$D^2 = \frac{1}{2}diag[C_1^2 \cdot I_{22}, \ldots, C_m^2 \cdot I_{22}] \tag{5.36}$$

and A is a perturbation which is norm bounded as,

$$\overline{\sigma}(\Delta) \leq \frac{m\pi C_{max}^2}{2N\nu} \tag{5.37}$$

Hence as the number of taps becomes large, the TDL confluence matrix (assuming the adaptive gain is normalized as $\mu = \overline{\mu}/N$), approaches the pairwise diagonal matrix D^2. As pointed out in Remark 3.1, this is precisely the form needed to ensure that the adaptive system is LTI. Hence for a large number of taps, the stability analysis reduces to examining pole locations of Laplace transforms. This was precisely Glover's approach in [9]. However, the need for the confluence matrix to be pairwise diagonal is somewhat restrictive (i.e., for the TDL case this requires an infinite number of taps), compared to the main results in the present paper which only require that the confluence matrice is positive definite.

4. The dependence of the condition number bound (5.19) on the inverse of $\underline{\nu}$ indicates that more taps N are required to "resolve" more closely spaced tones, This intuitive notion has been discussed heuristically in the literature (cf., Glover [9]). The analytical result (5.19) makes this dependence explicit.

5. The dependence of (5.19) on m indicates that more taps N are required to "resolve" more tones. This observation appears to be new.

6 CONCLUSIONS

The main results of the paper show that if the confluence matrix is positive definite the adaptive feedforward operator $\mathcal{H}$ from the error e to the estimate $\hat{y}$ is input-output equivalent to an adaptive system with a PE regressor. This implies that tracking error convergence is exponential for a large class of overparametrized adaptive feedforward systems. Furthermore, certain robustness properties of PE systems which are invariant under this internal reparametrization of $\mathcal{H}$ (e.g., such as boundedness with respect to additive noise disturbances in e) carry over to the overparametrized case.

These results generalize the results of Glover to a wider range of regressor basis functions, excitations, and adaptive algorithms. The results also emphasize the role of the confluence matrix (rather than the autocorrelation matrix) in determining the convergence properties of overparametrized adaptive systems.

It is concluded that the main penalty incurred from overparametrization is that the optimal rate of exponential tracking error convergence is degraded by the condition number of the confluence matrix. While this condition number can be quite large and lead to sluggish performance, it does not invalidate the overparametrization approach. On the contrary, now that it is known that overparametrized adaptive systems have exponential convergence just like full PE systems, they can be designed to have comparable performance if care is taken to ensure that the condition number of the confluence matrix is well behaved.

Properties of the confluence matrix were examined in detail for the special case of a Tap-Delay Line regressor with sinusoid excitation. A bound derived on the condition number of the confluence matrix indicates the effect of the number of taps, number of tones, tone spacing, delay time, etc. on the degradation of the optimal exponential convergence rate. Interestingly, as the number of taps increases, a limit is reached which depends on the ratio of magnitudes of largest and smallest tone amplitudes in the regressor. This is essentially the limiting penalty for TDL overparametrization with a large number of taps. The formula also indicates "diminishing returns" for adding taps beyond a certain point. This provides a simple rule of thumb for determining the number of taps in practice. In the limiting case where the number of taps goes to infinity, the confluence matrix becomes pairwise diagonal which implies that the adaptive system is LTI. This completely recovers Glover's results as a special asymptotic case of the present analysis.

Acknowledgements

This research was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

A APPENDIX A

LEMMA 1 *(Sastry and Bodson [14])* Consider the error equation,

$$e = \phi^T x \tag{A.1}$$

$$\dot{\phi} = -\mu x x^T \phi \tag{A.2}$$

where $\phi(t)$, $x(t) \in R^n$. Let x be a bounded piecewise continuous function of t such that,

$$||x(t)|| \leq \bar{x} < \infty; \text{ for all } t \geq 0 \tag{A.3}$$

and let there exist constants $\alpha_1, \alpha_2, \delta > 0$ such that the following PE condition is satisfied,

$$\alpha_1 I \leq \int_t^{t+\delta} x(t)x(t)^T dt \leq \alpha_2 I, \quad \text{for all } t \geq 0 \tag{A.4}$$

Then the system (A.1)(A.2) is *globally exponentially stable,* i.e.,

$$||\phi|| \leq p_0 e^{-\alpha t} \tag{A.5}$$

$$|e| \leq p_0 \bar{x} e^{-\alpha t} \tag{A.6}$$

where.

$$\alpha = \frac{1}{2\delta} \ln \left(\frac{1}{1-\alpha_3} \right) \tag{A.7}$$

$$p_0 = \left(\frac{1}{1-\alpha_3} \right)^{\frac{1}{2}} ||\phi(0)|| \tag{A.8}$$

$$\alpha_3 = \frac{2\mu\alpha_1}{(1+\mu\alpha_2\sqrt{n})^2} \tag{A.9}$$

Letting μ be sufficiently small (i.e., such that $\mu << 1/(\alpha_2\sqrt{n})$), gives,

$$\alpha \simeq \mu\alpha_1/\delta \tag{A.10}$$

$$p_0 \simeq (1 + \mu\alpha_1)||\phi(0)|| \tag{A11}$$

PROOF: The proof follows directly from the development in Sastry and Bodson [14] pg. 73-75 (see in particular Theorem 2.5.3) specialized to' the gradient adaptation algorithm (A.1)(A.2).

References

[1] K.J. Astrom and T. Bohlin, "Numerical identification of linear dynamic systems from normal operating records ," In *Theory of Self-Adaptive Control Systems.*, pp. 96-111, Plenum Press, New York, 1966.

[2] D.S. Bayard, J. Spanos and Z. Rahman, "Exponential convergence of the tracking error in adaptive systems without persistent excitation," IEEE Conf. on Decision and Control, New Orleans, LA, pp. 208-209, December 1995; Full version in JPL Internal Document, JPL D-12877, August 25, 1995.

[3] D.S. Bayard, *A General Theory of Linear Time Invariant Adaptive Feedforward Systems with Harmonic Regressors.* Jet Propulsion Laboratory, Internal Document JPL D-13556, April 12, 1996.

[4] D.S. Bayard, "Necessary and sufficient conditions for LTI representations of adaptive systems with sinusoidal regressors," Proc. American Control Conference, Albuquerque NM, June 1997.

[5] D.S. Bayard, "An LTI/LTV Decomposition of Adaptive Feedforward Systems with Sinusoidal Regressors," Proc. American Control Conference, Albuquerque NM, June 1997.

[6] D.S. Bayard, "LTI Representations of adaptive systems with tap delay-line regressors under sinusoidal excitation," Proc. American Control Conference, Albuquerque NM, June 1997.

[7] R.R. Bitmead, "Persistence of excitation conditions and the convergence of adaptive schemes," IEEE Trans. on Information Theory, vol. 30, pp. 183-191, March 1984.

[8] R.R. Bitmead and B.D.O. Anderson, "Lyapunov techniques for the exponential stability of linear difference equations with random coefficients," IEEE Trans. Automatic Control, vol. AC-25, pp. 782-787, August 1980.

[9] J.R. Glover, Jr., "Adaptive noise canceling applied to sinusoidal interferences," IEEE Trans. Acoustics, Speech and Signal Processing, vol. 25, no. 6, December 1977.

[10] G.H. Golub and C.F. Van Loan, *Matrix Computations.* 2nd Edition, Johns Hopkins University Press, Baltimore Maryland, 1989.

[11] P.A. Ioannou and J. Sun, *Robust Adaptive Control.* Prentice Hall, Upper Saddle River, New Jersey, 1996.

[12] R. Johansson, "Global Lyapunov stability and exponential convergence of direct adaptive control," Int. J. Control, vol. 50, no. 3, pp. 859-869, '1989.

[13] K.S. Narendra and A.M. Annaswamy, *Stable Adaptive Systems.* Prentice-Hall, Englewood Cliffs, New Jersey, 1989.

[14] S. Sastry and M. Bodson, *Adaptive Control: Stability, Convergence, and Robustness.* Prentice Hall, 1989.

[15] J. Spanos and Z. Rahman, "Narrow-band control experiments in active vibration isolation," SPIE Proc., Vibration Monitoring and Control, paper 2264-01, San Diego, July 1994.

[16] A. Weiss and D. Mitra, "Digital adaptive filters: Conditions for convergence, rates of convergence, effects of noise and errors arising from implementation," IEEE Trans. on Information Theory, vol. IT-25, pp. 637-652, November 1979.

[17] B. Widrow and S.D. Stearns, *Adaptive Signal Processing.* Prentice-Hall, Englewood Cliffs, New Jersey, 1985.

[18] J. S-C Yuan and W.M. Wonham, "Probing signals for model reference identification," IEEE Trans. Automatic Control, vol. 22, pp. 530-538, August 1977.